Antonis Maronikolakis, PhD Student

github email linkedin website scholar

Education

PhD Supervised by Hinrich Schütze — January 2020 - Current
Ludwig-Maximilians-Universität, Munich, Germany
Topic: Multilingual Hate Speech Detection and Analysis
Expected Graduation Date: 2023

MSc Speech and Language Processing - First Honours (1:1) — September 2018 - September 2019
University of Sheffield, Sheffield, UK
Thesis: Text Generation for News Headlines

BSc Computer Science - 8.4/10.0 (Upper 2:1) — October 2014 - June 2018
University of Piraeus, Athens, Greece
Thesis: Approximation and Heuristic Algorithms for Ridesharing

Publications

2022

Listening to Affected Communities to Define Extreme Speech: Dataset and Experiments. Antonis Maronikolakis, Axel Wisiorek, Leah Nann, Haris Jabbar, Sahana Udupa and Hinrich Schütze. 2022. In Findings of the Association for Computational Linguistics: ACL 2022.

Ethical Scaling for Content Moderation: Extreme Speech and the (In)Significance of Artificial Intelligence. Sahana Udupa, Antonis Maronikolakis, Hinrich Schütze, Axel Wisiorek. 2022. Harvard Kennedy School, Shorenstein Center.

Separating Hate Speech and Offensive Language Classes via Adversarial Debiasing Shuzhou Yuan, Antonis Maronikolakis, Hinrich Schütze. 2022. The 6th Workshop on Online Abuse and Harms (NAACL).

Analyzing Hate Speech Data along Racial, Gender and Intersectional Axes. Antonis Maronikolakis, Philip Baader, Hinrich Schütze. 2022. 4th Workshop on Gender Bias in Natural Language Processing (NAACL).

2021

Wine is Not v i n. – On the Compatibility of Tokenizations Across Languages. Antonis Maronikolakis, Philipp Dufter, and Hinrich Schütze. 2021. In Findings of the Association for Computational Linguistics: EMNLP 2021.

Artificial Intelligence, Extreme Speech, and the Challenges of Online Content Moderation. Sahana Udupa, Elonnai Hickok, Antonis Maronikolakis, Hinrich Schuetze, Laura Csuka, Axel Wisiorek, Leah Nann. 2021. AI4Dignity Project. EU Commission Policy Brief.

BERT Cannot Align Characters. Antonis Maronikolakis, Philipp Dufter, and Hinrich Schütze. 2021. In Proceedings of the Second Workshop on Insights from Negative Results in NLP (EMNLP).

Identifying Automatically Generated Headlines Using Transformers. Antonis Maronikolakis, Hinrich Schütze, and Mark Stevenson. 2021. In Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda (NAACL).

Multidomain Pretrained Language Models for Green NLP. Antonis Maronikolakis and Hinrich Schütze. 2021. In Proceedings of the Second Workshop on Domain Adaptation for NLP (EACL).

2020

Analyzing Political Parody in Social Media. Antonis Maronikolakis, Danae Sánchez Villegas, Daniel Preotiuc-Pietro, and Nikolaos Aletras. 2020. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

Professional Experience

» Junior Applied Scientist at Zalando
Customer Reviews Moderation and Ranking June 2022 - November 2022

- Worked in the domain of fashion customer reviews, researching methods to improve reviews moderation and ranking.
- Developed a few-shot learning method to improve over the production model for reviews moderation.
  - Finetuned BERT models using pattern-exploiting few-shot learning (PET), training both in monolingual and multilingual settings.
  - Showed that with multilingual transfer performance is improved in lower-resource settings.
  - Run a Monte Carlo simulation to identify training examples most conducive to performance.
  - With the proposed method, the production model was outperformed, using only a fraction of the training data (32 versus 20k).
- Explored methods to improve ranking of customer reviews, aiming to rank higher-information reviews closer to the top for better visibility. Tested methodology through a customer A/B test.
- Undertook onboarding and managing duties for newer members when my lead had to take an extended leave.

» NLP Researcher at AI4Dignity
Hate Speech Data Analysis for Marginalized Global Communities January 2021 - December 2021

- Worked in a multi-disciplinary team of NLP researchers, anthropologists and fact-checkers.
- Led the data collection and annotation team, culminating in a dataset of 20k examples and multiple label categories from a team of 13 annotators spanning 4 countries (Brazil, Germany, India and Kenya).
- Developed machine learning models, performed data analysis and deployed a proof-of-concept tool to combat hate speech in the examined countries.
- Our work and dataset is to be published in Findings of ACL 2022. As an extension of our work, we have also published a policy brief at the EU Commission.

» Speech Processing Practitioner for VoiceBase Research Lab
Analyzed Variational Autoencoders and Phrase Detection methods November 2018 - August 2019

- As part of Dr. Thomas Hain's VoiceBase research lab, I assisted researchers by exploring a novel Variational Autoencoder structure, alongside work on Phrase Detection and Extraction.
- Analyzed and ran experiments with the Factorized Hierarchical Variational Autoencoder repository and on various Phrase Detection algorithms.
- Set up and documented experiment configurations for ESPNet and DeepSpeech.

» AI Programmer/Writer for Google Summer of Code
Mentored under Dr. Peter Norvig, working on AI algorithms June - September 2017

- Worked on the Python repository of Dr. Norvig's book, *Artificial Intelligence: A Modern Approach.*
- Implemented, evaluated and wrote about Natural Language Processing and Machine Learning algorithms and concepts.
- Collaborated with Dr. Norvig to polish pseudocode from the book.
- After the end of the project, I was given administrator rights to the repository.

Tools & Technologies

Python, HuggingFace Ecosystem, LaTeX, Bash, AWS Ecosystem

Awards

- Scholarship for academic excellence - 2016-2017, 2017-2018
- Kaggle Machine Learning Competitions (2018) - Silver and Bronze Medals
- Microsoft Imagine Cup Competition (2015 and 2016) - National Finalist
- National Computer Science Competition 2014 - Finalist